Future Computational Needs for Climate Change Modeling

Michael Wehner

Lawrence Berkeley National Laboratory

mfwehner@lbl.gov

510-495-2527

Background

- Modeling climate change is high profile science
  - Politically and socially relevant
  - Often covered in the popular press
- Scientific progress is limited by observational and computational constraints.
- The complexity of the natural climate system must be reflected in numerical models if we hope to gain understanding.

General Circulation Models (GCMs)- Present

- Fully coupled GCMs are our most sophisticated computational tools.

- The current state of the art includes submodels of:
  - Atmosphere
  - Ocean
  - Sea ice
  - Land processes

- Essentially a fancy hydrodynamics model with thermodynamic source terms.
  - Tracks energy, momentum and moisture.

General Circulation Models (GCMs)-Future

- Current developmental models also target chemical processes.
- The most important of these from a climate change perspective is biological.
  - Add carbon and other nutrients to the prognosis
- Atmospheric chemistry adds numerous other prognostic variables.
- All of this will further add to the computational burden.

General Circulation Models (GCMs) Limitations

- GCMs are expensive.
  - They exhibit poor scalability in several respects.
- Example: Atmospheric dynamics
  - Stable finite difference solution of the Navier-Stokes equations is limited by the Courant Condition.
    - *?t < ?x/v*
    - *?x* = grid spacing, *v* = maximum wind speed
  - The consequence of this condition is that computational burden increases nonlinearly as resolution is increased.

General Circulation Models (GCMs) Limitations

- Climate change integrations must integrate for multiple centuries.
- Courant conditions are measured in minutes.
- This large number of time steps is the principal limiting factor in climate modeling.
  - Prevents the exploitation of large numbers of processors.
- Example.
  - Domain decomposition of grid based hydrodynamics schemes requires that the ratio of “interior” cells to “border” cells be high for parallel efficiency
  - Doubling the horizontal resolution = Four times as many cells
  - Four times as many processors can be used at the same efficiency
  - But each processor has twice as much work to do. Time step was halved. Run time is twice as long.

Community Climate System Model (NCAR)

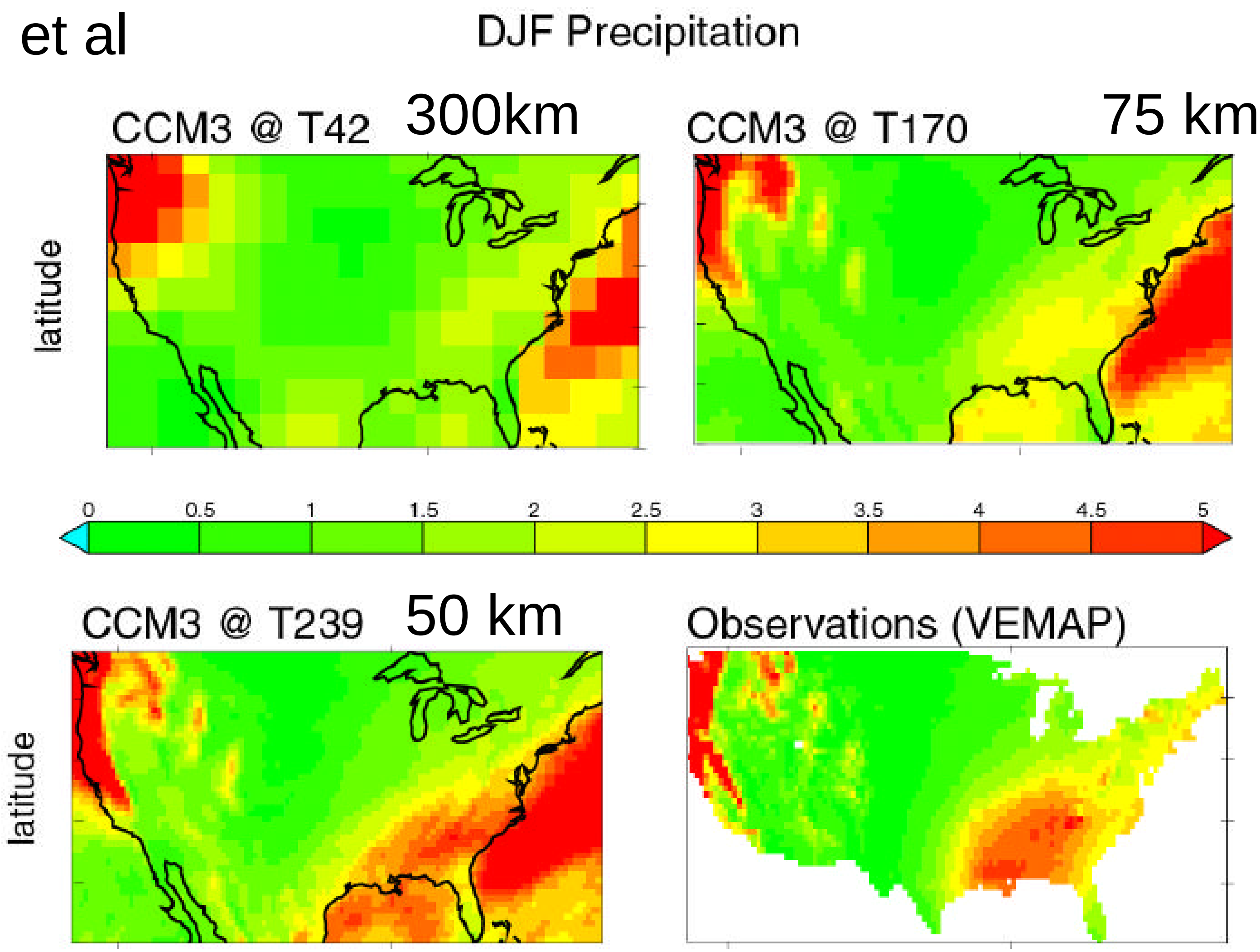
- Current resolution
  - Atmosphere: Grid=300km. ?t=20 minutes
  - Ocean: Grid=60km ?t=60 minutes
  - one simulated year takes five machine hours (128PE SP3)
  - 1000 year control run integration required 9 calendar months (7 charged months)
- Sounds bad, but this is really good!
  - 1990 AMIP1: Many modeling groups required a calendar year to complete a 10 year integration of a stand alone atmospheric general circulation model. Typical resolution was 600km or more.

What we really want.

- Atmosphere
  - Regional climate change prediction will require horizontal grid resolution of 10km (3600X1800)
  - Cloud physics parameterizations could exploit 100 vertical layers
- Ocean
  - Mesoscale (~50km) eddies are thought to be crucial to ocean heat transport
  - 0.1º grid will resolve these eddies (3600X1800)
  - Short stand-alone integrations are underway now.
- Ensembles of integrations are required to address issues of internal (chaotic) variability.
  - Current practice is to make 4 realizations. 10 is better.
- Time steps will be measured in seconds.

Simulated precipitation as a function of resolution

Duffy, et al



What kind of machine do we need?

- What we have is:
  - seaborg.nersc.gov: 6080 processor IBM Power 3
  - cheetah.ornl.gov: 864 processor IBM Power4
  - bluesky.ucar.edu 1024 processor IBM Power4
  - blackforest.ucar.edu: 1172 processor IBM Power 3
- Typical CCSM2 configuration uses 128 processors
  - 32 atmosphere (OpenMP and MPI)
  - 48 ocean (MPI only)
  - 8 land (OpenMP and MPI)
  - 32 ice (MPI only)
  - 8 coupler (OpenMP only)
- No significant speedup for additional processors.
- Achieves about 5% of peak on Power 3

What kind of machine do we need?

- Climate codes are highly vectorizable.
  - High resolution atmospheric portion achieves 22% of peak on 8 processors of NEC SX6 (14GFlops)
  - Turnaround is 3 to 6 times faster than the best I can do on Power3.

What kind of machine do we need?

- Two types of runs:
  - A single 1000 year control run to estimate noise.
  - Ensemble of 4-10 transient runs to estimate signal.
    - I.e. simulation of historical record 1870-2002 or a prediction of the future climate.
    - This is an additional embarrassingly parallel dimension.

What kind of machine do we need?

- Seaborg is the wrong kind of machine.
  - queue is tuned to large node jobs.
    - Justifiably so due to large investment in the network.
    - ensemble runs are made individually. Jobs are still small despite overall usage of more nodes.
  - 5% of peak?
    - Too slow to exploit large numbers of processors.

What kind of machine do we need?

- A machine with faster processors allows the usage of more processors for a single realization. Remember the courant condition.
  - Allows higher resolution with more processors at comparable parallel efficiency (grid cells per PE)
- Example: Increase atmospheric resolution from 300km to 10km.
  - If you increase the number of processors by 30^2 (900) to maintain the same parallel efficiency (~200 cells per processor) and
  - If you increase sustained per processor performance by a factor of 30 to compensate for the time step reduction then
  - ***Turnaround time stays the same***.
  - In other words, ~30000 processors each achieving ~2GFlops sustained. Six Earth Simulators.

Conclusions

- Despite large strides in computing ability, climate change prediction needs are not yet being met.
- Limitations are imposed by per processor sustained speed.
- Machines large numbers of processors could be exploited better but only when processors get much faster.
  - Large numbers of processors are not bad. Slow processors are bad.